

**PREPROCESSING DENGAN ITERATIVE PARTITIONING FILTER
ALGORITHM UNTUK KLASIFIKASI**

TUGAS AKHIR

**Diajukan untuk Memenuhi Sebagian Persyaratan Mencapai
Derajat Sarjana Teknik Informatika**



Disusun oleh :

Felik Ade Kristiawan

12 07 06836

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA
2016**

HALAMAN PENGESAHAN

Tugas Akhir berjudul
***PREPROCESSING DENGAN ITERATIVE PARTITIONING FILTER
ALGORITHM UNTUK KLASIFIKASI***

Disusun Oleh

Felik Ade Kristiawan

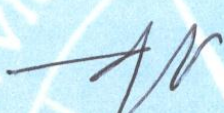

12 07 06836

Dinyatakan telah lengkap memenuhi syarat pada tanggal

22 Juli 2016

Dosen Pembimbing 1,

Dosen Pembimbing 2,

 Ir.A. Djoko Budiyan, M.Eng, Ph.D.  Dra. Ernawati, M.T.

Tim Penguji,

Penguji 1,

 Ir.A. Djoko Budiyan, M.Eng, Ph.D.

Penguji 2,

Penguji 3,

 Patricia Ardanari S.Si, M.T.

 Eddy Julianto S.T, M.T.

Yogyakarta, 22 Juli 2016

Universitas Atma Jaya Yogyakarta

Fakultas Teknologi Industri

 Dekan:

Dr. Teguh Siswanto
FAKULTAS TEKNOLOGI INDUSTRI

KATA PENGANTAR

Puji syukur kepada Tuhan atas semua karunia dan berkat-Nya yang telah dilimpahkan kepada penulis, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan baik. Tugas akhir ini adalah untuk memenuhi sebagian persyaratan mencapai derajat sarjana di Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.

Penulis menyadari bahwa dalam pembuatan skripsi ini tidak terlepas dari bantuan berbagai pihak yang telah menyumbangkan segala dukungan, tenaga, pikiran dan doa bagi penulis, sehingga pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Tuhan Yesus Kristus yang telah menyertai, memberi arahan dan memberkati penulis sehingga laporan tugas akhir ini dapat terselesaikan dengan baik.
2. Bapak Dr. A. Teguh Siswanto selaku Dekan Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
3. Bapak Martinus Maslim, S.T.,M.T., selaku Ketua Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
4. Bapak Ir.A Djoko Budiyo, M.Eng, Ph.D., selaku Dosen Pembimbing I yang telah meluangkan waktu dan pikiran untuk memberi bimbingan dan arahan bagi penulis dalam menyelesaikan tugas akhir ini.

5. Ibu Dra. Ernawati, M.T., selaku Dosen Pembimbing II yang telah meluangkan waktu dan pikiran untuk memberi bimbingan dan arahan bagi penulis dalam menyelesaikan tugas akhir ini.
6. Keluarga besar Kantor Tata Usaha Fakultas Teknologi Industri Universitas Atma Jaya, yang telah memberikan begitu banyak pengalaman, semangat dan dorongan bagi penulis, terimakasih telah menjadi rumah dan menjadikan penulis anak paling kecil di kantor ini selama penulis menjadi student staff.
7. Seluruh Dosen dan Staff Pengajar Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta yang telah membantu penulis selama masa kuliah di Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Atma Jaya Yogyakarta.
8. Tim Futsal FTI Universitas Atma Jaya Yogyakarta, Terima kasih untuk Pak Martinus S.T,M.T. berkat beliau penulis menjadi rajin berolahraga futsal secara gratis.
9. Seluruh keluarga, Bapak, Mama, Mbak Nana, Mas Agung, Mbak Valen, Mas Andri, Geby, Arya, Tyasmalinda sekeluarga dan keluarga besar yang selalu memotivasi, memberi doa, dan memberi semangat dengan berbagai cara dan memberikan kehangatan tersendiri bagi penulis sehingga penulis dapat menyelesaikan kuliah dan Tugas akhir ini dengan baik.

10. Keluargaku Komunitas Garuda Katolik yang telah menjadikan penulis seperti sekarang, berkat komunitas Garuda Katolik Penulis dapat menyirami iman penulis, dan menemukan teman baru dari berbagai jurusan, asal dan budaya, sehingga penulis dapat berbagi kisah suka duka dan tawa lepas tanpa malu.
11. Sahabat Four Idiots, Yoshua, Rivan, Agung, Terima kasih atas pengalaman berharganya selama ini, sukses untuk kalian kedepannya.
12. Teman-teman angkatan 2012 yang masih berjuang maupun yang sudah lulus terimakasih atas kerjasama dan dukungannya selama penulis melaksanakan studi di Universitas ini.
13. Teman-teman skripsi seperjuangan Beny, Pandhu, Hana eka, Reni, Raymond, Debora Dan Deo yang sudah lulus. Terima kasih telah menemani dalam berjuang bersama mengerjakan skripsi, sehingga skripsi ini dapat terselesaikan dengan baik.
14. Rico dan Hugo, terimakasih atas kerja samanya selama menjadi partner kerja praktek di PT.Medion Farma Jaya Bandung.
15. Anissa, Yeni, Jourgi, Tity, terimakasih telah menjadi partner sekaligus sahabat yang baik selama menjadi asisten dosen praktikum aplikasi komputer.
16. Matheus alvian, dan teman-teman hotel biru yang telah membantu mencairkan suasana selama penulis mengerjakan skripsi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki penulis. Oleh karena itu kritik dan saran yang membangun dari semua pihak sangat diharapkan untuk penyempurnaan yang akan datang. Akhir kata, penulis berharap supaya tugas akhir ini dapat bermanfaat bagi semua pihak.

Yogyakarta, Juni 2016

Penulis

Felik Ade Kristiawan

HALAMAN PERSEMBAHAN

Skripsi ini aku persembahkan untuk Tuhan Yesus Kristus pelindungku, Keluargaku, Pacarku, Sahabatku dan teman-teman dekatku.

MOTTO

..Dan kamulah satu satunya yang membuat duniaku menjadi lebih baik.. ~ You Are The Apple Of My Eye

“ Karena itu Aku berkata kepadamu : apa saja yang kamu minta dan doakan, percayalah bahwa kamu telah menerimanya maka hal itu akan diberikan kepadamu ”

Markus 11 : 24

NO PAIN NO GAIN ~ NO LIMITS

DAFTAR ISI

HALAMAN COVER.....	i
HALAMAN PENGESAHAN.....	ii
KATA PENGANTAR.....	iii
HALAMAN PERSEMBAHAN.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xiii
INTISARI.....	xv
BAB I PENDAHULUAN	
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	4
1.5. Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	
2.1 Tinjauan Pustaka.....	6
BAB III LANDASAN TEORI	
3.1 Data, Informasi, Pengetahuan.....	10
3.2 Data Mining.....	10
3.3 Preprocessing Data.....	12
3.4 Noisy Data filtering.....	13
3.5 Kompleksitas Data.....	14
3.6 Iterative Partitioning Filter.....	16
3.7 Jaringan Syaraf Tiruan.....	18
3.8 Algoritma Backpropagation.....	20
3.9 Knowledge Exctraction Based on Evolutionary Learning (KEEL).....	20
3.10 Waikato Environment for Knowledge Analysis (WEKA).....	21

BAB IV METODOLOGI PENELITIAN

4.1	Diagram Penelitian.....	24
4.2	Dataset Penelitian.....	25
4.3	Preprocessing.....	27
4.4	Klasifikasi.....	31
4.5	Evaluasi.....	32

BAB V HASIL DAN PEMBAHASAN

5.1	Potongan dataset sebelum IPF dan sesudah IPF.....	35
5.2	Preprocessing dengan Iterative Partitioning Filter dan Hasil Klasifikasi.....	36

BAB VI KESIMPULAN DAN SARAN

6.1	Kesimpulan.....	104
6.2	Saran.....	105

DAFTAR PUSTAKA.....	106
---------------------	-----

DAFTAR GAMBAR

Gambar 3.1 <i>Non-overlapping and Overlapping area of cluster c_1 and c_2</i>	15
Gambar 3.2 Diagram Algoritma Iterative Partitioning Filter.....	17
Gambar 3.3 Arsitektur Multi Layer Perceptron (Fahlman & Hinton, 1987).....	19
Gambar 3.4 Tampilan awal keel.....	21
Gambar 3.5 Tampilan awal Weka.....	22
Gambar 4.1 Aliran Diagram Penelitian.....	24
Gambar 4.2 Pengaturan Parameter Iterative Partitioning Filter.....	28
Gambar 4.3 Perancangan Framework Preprocessing Eksperimen Keel.....	30
Gambar 4.4 pengaturan parameter Multilayer Perceptron Backpropagation.....	32
Gambar 5.1 Potongan dataset Iris sebelum preprocessing dan sesudah di preprocessing dengan IPF.....	35
Gambar 5.2 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Wine dengan skema majority dan nilai confidence 0.25, 0.50, 0.75, dan 1.....	41
Gambar 5.3 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Wine dengan skema consensus dan nilai confidence 0.25, 0.50, 0.75 dan 100.....	48

Gambar 5.4 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Iris dengan skema Majority dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	54
Gambar 5.5 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Iris dengan skema consensus dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	61
Gambar 5.6 Grafik perbandingan waktu pelatihan terhadap number partitions.....	61
Gambar 5.7 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Wisconsin dengan skema Majority dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	67
Gambar 5.8 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Wisconsin dengan skema Consensus dan nilai confidence 0.25, 0.50,0.75 , dan 1.....	72
Gambar 5.9 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Pima dengan skema Majority dan nilai confidence 0.2, 0.50, 0.75, 1.....	77
Gambar 5.10 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset Pima dengan skema consensus dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	82
Gambar 5.11 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan	

dataset haberman dengan skema majority dan nilai confidence 0.25,0.50, 0.75 & 1.....	87
Gambar 5.12 Grafik perbandingan hasil klasifikasi dengan parameter NumberPartitions pada percobaan dataset haberman dengan skema consensus dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	93
Gambar 5.13 Grafik Perbandingan hasil akurasi terbaik masing-masing dataset terhadap tipe filter.....	96
Gambar 5.14 Kurva ROC Wine Kelas 1.....	98
Gambar 5.15 Kurva ROC Wine Kelas 2.....	98
Gambar 5.16 Kurva ROC Wine Kelas 3.....	99
Gambar 5.17 Kurva ROC Iris Kelas Iris-Versi Color....	99
Gambar 5.18 Kurva ROC Iris Kelas Iris-Setosa.....	100
Gambar 5.19 Kurva ROC Iris Kelas Iris-Virginica.....	100
Gambar 5.20 Kurva ROC Wisconsin Kelas 2.....	101
Gambar 5.21 Kurva ROC Wisconsin Kelas 4.....	101
Gambar 5.22 Kurva ROC Pima Kelas Tested_Positive....	102
Gambar 5.23 Kurva ROC Pima Kelas Tested_Negative....	102
Gambar 5.24 Kurva ROC Haberman Kelas Positif.....	103
Gambar 5.25 Kurva ROC Haberman Kelas Negatif.....	103

DAFTAR TABEL

Tabel 3.1. Karakteristik dataset dan Pengukuran Kompleksitas untuk Pengujian Overlap Clustering.....	16
Tabel 4.1 Spesifikasi Pengukuran Hasil Eksperimen....	33
Tabel 4.2 Confusion Matrix klasifikasi.....	34
Tabel 5.1 Hasil eksperimen dataset Wine dengan skema majority, dan nilai confidence 0.25 & 0.50.....	36
Tabel 5.2 Hasil eksperimen dataset Wine dengan skema Consensus, dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	43
Tabel 5.3 Hasil eksperimen dataset Iris dengan skema Majority, dan nilai confidence 0.25,0.50,0.75 dan 1.....	49
Tabel 5.4 Hasil eksperimen dataset iris dengan skema Consensus, dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	56
Tabel 5.5 Hasil eksperimen dataset Wisconsin dengan skema Majority, dan nilai confidence 0.25, 0.50, 0.75, 1.....	63
Tabel 5.6 Hasil eksperimen dataset Wisconsin dengan skema Consensus, dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	68
Tabel 5.7 Hasil eksperimen dataset pima dengan skema Majority, dan nilai confidence 0.25, 0.50, 0.75, dan 1.....	73

Tabel 5.8 Hasil eksperimen dataset pima dengan skema Consensus, dan nilai confidence 0.25,0.50,0.75 dan 1.....	78
Tabel 5.9 Hasil eksperimen dataset haberman dengan skema majority, dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	83
Tabel 5.10 Hasil eksperimen dataset haberman dengan skema consensus, dan nilai confidence 0.25, 0.50, 0.75 dan 1.....	89
Tabel 5.11 Hasil perbandingan skema filter untuk akurasi terbaik dengan numberpartitions minimal.	96
Tabel 5.12 Hasil perbandingan masing-masing dataset sebelum di preprocessing dan setelah di preprocessing dengan dua buah skema penyaringan Iterative Partitioning Filter.....	97

INTISARI

Preprocessing data dan analisis kinerja *preprocessing* data sangat penting dalam *data mining*. Kedua hal tersebut memiliki dampak besar pada keberhasilan proses *data mining*, karena keputusan-keputusan yang berkualitas harus didasarkan pada data yang berkualitas. Dengan *preprocessing* maka dapat membantu kita untuk meningkatkan kualitas data dan menghapus *noise* data.

Pada penelitian ini, dibahas bagaimana kinerja dari algoritma *noisy data filtering* yaitu *iterative partitioning filter* dengan menggunakan berbagai dataset dari University of California, Irvine (UCI) Machine Learning Repository dengan variasi parameter *iterative partitioning filter* yang dirubah. Penelitian ini juga membahas bagaimana menganalisis hasil klasifikasi dari dataset yang telah di *preprocessing* menggunakan *Backpropagation* sehingga dapat mengidentifikasi akurasi terbaik dari berbagai dataset yang telah diuji.

Hasil akhir dari penelitian ini berupa data tabel hasil eksperimen yang terdiri dari waktu pelatihan, akurasi klasifikasi, Kesalahan klasifikasi, *Kappa statistic*, *Mean Absolute Error (MAE)* atau rata-rata error periterasi sesuai dengan data, *Root mean squared error* dan *confusion matrix*. Hasil akhir yang ditampilkan juga berupa grafik hasil eksperimen yang dibandingkan dengan parameter *iterative partitioning filter* yang divariasi.

Kata Kunci : *Data mining, Iterative Partitioning Filter, Backpropagation, UCI Machine Learning Repository.*